DOCUMENT RESUME

ED 023 410 52 LI 000 370

On Library Statistics. Submitted to the National Advisory Commission on Libraries.
Mathematica, Princeton, N.J.; National Advisory Commission on Libraries, Washington, D.C.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-7-0961
Pub Date Aug 67
Contract-OEC-2-7-010105-1523
Note-71p.
EDRS Price MF-$0.50 HC-$3.65
Descriptors-College Libraries, Data Collection, *Libraries, Literature Reviews, Measurement Techniques, Research Methodology, Sampling, School Libraries, Special Libraries, Statistical Analysis, *Statistical Surveys, *Statistics, University Libraries

Available library statistical information is unreliable and missing in many areas. This report covers three general subjects: (1) the description of the nature of the figures that are currently available, including lists of library statistics publications, and an indication of some information which is wanted but not available; (2) a discussion of the philosophy which might appropriately underlie a systematic data collection effort; and (3) some suggestions for the organization of data collection activities. Library statistical data serves primarily four major classes of users, (1) the librarian, (2) those persons who deal professionally with inter-library matters, (3) instructional users of library services, and (4) those who provide the funds upon which libraries operate. Statistics, in order to be useful, must potentially lead to a modified course of action. Thus, data should be provided in a conditional form associating them with values of related variables, and a list made of alternatives needed. Major recommendations made to the National Advisory Commission on Libraries are: (1) consideration be given to the publication of an annual compendium of library statistics, (2) data to be collected, whenever possible, on the basis of a constant sample, (3) data be collected which is necessary for the planning of capital construction. Suggestions are also made for improving the quality of statistics for each type of library. (CM)

BR-7-0941
P.A.5
LI 000 370
ERIC/CLIS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED023410

LI 000 370

MATHEMATICA
One Palmer Square
Princeton, New Jersey

ON LIBRARY STATISTICS

August 1967

Submitted to

NATIONAL ADVISORY COMMISSION ON LIBRARIES

ON LIBRARY STATISTICS

Part I. Introduction [1]

We have rarely encountered an area of data collection which approximates the libraries in terms of the quality of thought that has been devoted to the mapping of the terrain and the small quantity of systematic information that has actually been accumulated. The field has been provided with elaborate reports specifying in considerable detail the types of statistical series which it would be desirable to assemble, the definitions of the variables which might most profitably be employed and the likelihood that one can in fact obtain, in reliable and readily interpretable form, each type of information specified. These studies are impressive in that they appear to combine a degree of understanding of library operation which can only be attained by a professional librarian with an unusual degree of sophistication in statistical matters.

Yet, in our work on the economics of library operation we found that the available statistical data might be described with little exaggeration as a collection of gaps interspersed by an occasional bit of reliable information. Where statistical series seem to be available, closer examination showed that the reported figures were unreliable, and were sometimes put together in a way

[1] We would like to express our deep appreciation to the various persons at the U.S. Office of Education, whose help in providing material for the study was absolutely essential for its completion.

that rendered them totally unusable. Moreover, it is by no means clear that matters have been progressing in this area. For example, while it is impossible to determine just how much federal money is devoted to the collection and processing of library statistics, since so many of these costs are inextricably interwoven with other types of outlay, someone in a position to make a reasonable surmise has ventured the view that governmental expenditures for this purpose have, on balance, been declining. We do know that some important statistical series which were available in the preceding decade have been discontinued. All of this suggests that there are grounds for uneasiness in the current state of library statistics and that a reexamination of the subject, including a status report, is entirely appropriate.

This report undertakes to treat three general subjects. First, we describe in some detail the nature of the figures that are currently available and indicate some of the information that has been called for by the literature but which is not in fact available. Second, there is a discussion of what might be called the meta-strategy of data collection, as it were, the philosophy which might appropriately underlie a systematic data collection effort and its implications for the specific features of such an undertaking. Finally, we will make some specific suggestions for the organization

of data collection activities which may perhaps prove helpful in their implementation. These recommendations may possibly point the way toward a more systematic and effective program for the accumulation of the data necessary for rational planning and decision making in matters relating to the nation's library resources.

Part II. Survey of the Current State of the Statistics

A. Public School Libraries

The data relating to Public School Libraries are perhaps the poorest among the available library statistics. A series of periodic general surveys entitled "Statistics of Public School Libraries" was published in the Biennial Survey of Education by the Office of Education for the fiscal years 1935, 1942, 1948 and 1954.[2] As was the case for college and university, and public libraries most of the data (for 1935, 1942, 1948) purporting to provide the aggregate U.S. figures were obtained by adding together the figures in the questionnaires that happened to be returned, and omissions consequently were misrepresented as zeros.

The 1954 figures, with a later companion publication[3] seem to provide a more accurate set of data. The statistics are

---

[2] Of even more questionnable quality is a publication put out in 1926 by the Bureau of Education, Statistics of Public Society, and School Libraries.

[3] U.S. Office of Education, Washington, D.C., Statistics of Public School Libraries 1960-61 Part I Basic Tables, Part II Analysis and Interpretation.

sometimes reported for all public libraries together, while at other times they refer to school districts enrolling 150 pupils or more and/or districts with centralized libraries.

This variation in the sample can constitute a frustrating inconvenience. For example, in the 1961 data, salary expenditure figures are only provided for school districts enrolling 150 pupils and over, but a breakdown for "other expenditures" (materials, binding and supplies) exists only for all public schools together. It would seem more profitable to provide all figures in terms of one basic sample, accompanied by a few comparisons with figures for the total universe of American public school libraries to illustrate the sample's relative size.

Two shorter supplementary bulletins of Public School Library Statistics were published by the Office of Education for the fiscal years 1959 and 1963. They were not meant to be part of a series of comprehensive studies, but were intended to "provide basic data within a short time." [4] They seem to be accurate and helpful, though the data are restricted to schools with centralized libraries in school systems with 150 pupils or more.

---

[4] Statistics of Public School Libraries 1960-61, Part I, p. iii.

Hence, for the moment, no consistent time series can be constructed from the data available for public school libraries. It may be noted before turning from this subject, that any systematic data collection effort in this area will require a choice between the collection of figures for all types of public school libraries together and their separation in terms of elementary, junior high, and high school libraries. Current trends suggest that the distinction among the three levels of schooling will probably become increasingly blurred and this will perhaps necessitate aggregation of the statistics.

B. College and University Libraries.

"Since 1870 the Office of Education has conducted statistical surveys of libraries. These appeared originally as 'Statistics of Public, Society and School Libraries.' As the number of libraries grew and their functions and activities became more specialized, separate studies of the various types of libraries were released. During the 1930's a new pattern of library statistics was established which resulted in separate publications on public, college and university, and public school libraries."[5]

---

[5] Office of Education, Library Services Branch, Statistics of Public Library Systems Serving Populations of 50,000 to 99,999: fiscal year 1960. March, 1962 (OE-15034), p. 1.

The earliest data for a cross-section of U.S. college and university libraries appear in the Biennial Survey of Education in the United States, "College and University Library Statistics," U.S. Office of Education for the fiscal years 1940, 1947 and 1952. The data give grand totals for all libraries in the continental United States, with a breakdown by type of institution. There are, however, two severe drawbacks to this set of data which render them virtually useless.

a) No adjustments were made for non-respondents. Questionnaires were sent to the president of every college and university listed in the Educational Directory of the Office of Education. The figures in the returns were simply summed and the totals were taken to represent the aggregate data for all of the United States.

b) In calculating average figures, whenever a library did not report an item, the omission was averaged in as a zero. The number of libraries reporting is noted, but failure to respond to a specific question is not recorded. Hence in a year when, say, the figure for average librarian salary is relatively low, it is literally impossible to tell whether this reflects a reduction in the level of payments on an unusually large number of non-respondents to the question. In 1957, a transitional report was

issued by Office of Education which consisted of a sample of "Larger Colleges and Universities."

The longest and most reliable set of data was initiated by the Statistics Committee of the Association of College and Research Libraries of the American Library Association, and provided by them from 1929 to 1960. These studies, covering a 35 to 50% sample of all academic libraries, appeared annually in the January issues of College and Research Libraries. The data prior to 1951 were collected on a limited scale. From 1951 to 1960 the data represent a comprehensive institutional coverage of academic libraries. Unfortunately, until 1960, aggregates were not calculated from the sample. Hence, unless one is interested in a single or small group of institutions, or is willing to compute his own sample aggregates, (as was found necessary to do in our study on the economics of library operation), the institutional data remain unilluminating because of their unwieldly detail.

This survey was continued from 1960 to 1964 by the U.S. Department of Health, Education, and Welfare, Office of Education, as Library Statistics of Colleges and Universities, Institutional Data. An analytic report for fiscal 1960 was issued in July 1961. An analytic report also accompanied the 1962 study. There also exist two unpublished Analytic Reports for 1963 and 1964 as well as

two specialized ALA Institutional supplements for the same years.

The 1960 Analytic Report gives median figures for the aggregate United States. The 1962 Analytic Report gives totals (for both 1960 and 1962) as well as frequency distributions for the aggregate U.S. The data collected for fiscal 1965 are currently being processed.

In 1966 the American Library Association assumed responsibility for the library surveys, and published, Library Statistics of Colleges and Universities 1965-66 Institutional Data. The study not only includes institutional data for almost 1,900 academic libraries, but also gives grand totals for the aggregate U.S. for 1960 to 1966.

This sixteen year time series of institutional data for college and university libraries is, to the best of our knowledge, the longest and most complete set of figures available for any of the various types of libraries (public, public school, state or specialized). As described in the last section of this report, however, further revision can increase its value even further.

C. Public Libraries

For the fiscal years 1939, 1945 and 1950 the U.S. Office of Education issued Bulletins entitled Public Library Statistics. Questionnaries were sent out to all Public Libraries in the United States. The figures reported in the returns were simply summed

together to form a total U.S. figure. Thus institutions which did not report, say, total operating expenditures, would in effect have been represented to have incurred no operating expenditure. Adjustments designed to avoid this error were undertaken in The Statistics of Public Libraries 1955-56, published by the Office of Education in the Biennial Survey of Education in the United States, 1954-56. For the first time in this series a sampling technique was tried.

The next comprehensive report in this series (for fiscal 1962) appeared in two parts: one dealing with libraries serving populations no smaller than 35,000, and one based on a study by the University of Illinois of libraries serving smaller populations.[6] Because of the almost impossible task of collecting a representative sample of data from the nation's numerous small libraries,[7] the second of these reports is not generally considered to constitute an accurate survey.

---

[6] U.S. Office of Education, Statistics of Public Libraries Part I Selected Statistics of Public Libraries Serving Populations of 35,000 and above. Institutional Data; the federal government sponsored a study by the University of Illinois, Graduate School of Library Science, 1962 Statistics of Public Libraries Serving Populations of less than 35,000.

[7] The Office of Education estimates that in 1962 almost 70% of the 7,257 public libraries served a population of less than 10,000 persons.

In order to provide information more frequently, in 1945 the Office of Education began to issue annually a limited amount of basic data.

"Originally this series covered two groups of cities, those with populations of 50,000 to 99,999 and those with 100,000 or more. Since then, three other publications for county and regional libraries and those serving smaller population groups were added. In 1960, the Library Services Branch issued five separate publications of 1959 data, one for public libraries in cities with populations of 100,000 or more (OE-15014A Rev.), one for those in cities of 35,000 to 49,999 (OE-15016), one for county and regional public library systems serving populations of 50,000 or more (OE-15017), and one for county and regional public library systems serving 35,000 to 49,999 (OE-15018). For 1960, the data for these five population groups were merged in three publications which cover public libraries whether municipal, county, or regional serving populations of

(1) 100,000 or more (OE-15033)

(2) 50,000 to 99,999 (OE-15034)

(3) 35,000 to 49,999 (OE-15035)"[8]

---

[8] Office of Education Library Services Branch. Statistics of Public Library Systems Serving Populations of 50,000 to 99,999: Fiscal Year 1960 (OE-15034) p. 1, March, 1962.

The following table lists all of these studies which exist to the best of our knowledge.

Population Breakdown (in 000's)	Type of Area Served	Years for which Series Appears
35 - 50	County/Regional	'59
50 +	County/Regional	'45 (no totals); '52 (only a sample); '53,'54,'55,'56,'57,'58,'59
35 - 50	City	'57,'58,'59
50 - 100	City	'45 (no totals) limited ed. in 1950, '52,'53,'54,'55,'56,'57,'58,'59
100 +	City	'45,'46,'47,'48,'49,'50,'51,'52,'53,'54,'55,'56,'57,'58,'59
25 - 35	City	'45 (no totals)
35 - 50	Total	'60 (includes also a '59 table) '62
50 - 100	Total	'60 (includes also a '59 table) '62
100 +	Total	'60 (includes also a '59 table) '60 and '60 (including N.Y. reference library)
35 and Below	Total	'62 (questionable set of dat

The earlier publications are much more limited in scope than the later ones. Many of the figures (in particular the expenditure statistics) for years before 1954 were later revised to "make them comparable" to the more recent data.

So far, no really complete and current time series can be constructed from these surveys. For our report on the economics of libraries, we chose figures for the years 1954 to 1959, for libraries serving a population of 50,000 or more, as the most consistent set of data available from the public library statistics. Happily, the revised fiscal 1960 and 1962 reports suggest there will eventually be available a consistent, comprehensive series of data.

D. Other Types of Library

Data on special and state libraries are extremely spotty. There exist isolated reports such as Survey of Special Libraries Serving State Governments, 1965, by Robert Galvik (to be released in 1967 by the U.S. Office of Education). There are no previous reports on special libraries by the Office of Education. In preparation by the Federal Library Committee, the Library of Congress and the Office of Education is a Survey of Special Libraries Serving the Federal Government 1965-66 by Paul Howard (to be published during 1967). A few infrequent publications have discussed Health Science, Medical and Law Libraries. However, there is simply no data series tracing developments relating to special and state library statistics.

E. Types of Information Currently Provided

Having completed our review of the sorts of information provided for the various types of library, it is convenient to provide

an overall summary indicating the classes of information available for libraries in the various categories. Table I illustrates some of the gaps in the data collection for the nation's libraries. The first column specifies some of the categories of statistics most often called for in the literature or most obviously needed. Each other column represents the most current reliable report available for the corresponding type of libraries. A "yes" indicates that the specified information appears in that report, a "dash" indicates that it is not included. (Sometimes the necessary data have been collected, but were not culled from the remaining statistics).

The table places library statistics in a more favorable light than they deserve. Each of the reports representing public school libraries, and public libraries is comparable to only one other corresponding set of figures (for 1956 and 1960 respectively). The 1966 data for college and university libraries exist in comparable form for 1960 through 1966. Earlier institutional figures (for 1951 through 1960) do exist, but they were never aggregated. Should they be totalled, the result would be a time series for 1951 through 1966 for college and university libraries.

State and special libraries are not even entered on the table, since there is no comprehensive set of figures which could represent

Table I. Types of Statistical Information Available by Class of Library [9]

	Public Libraries (Serving populations of 35,000 and above) (1962)	College and University Libraries (1966)	Public School Libraries (1961)	
Number of Libraries (Number in Sample)	Approximate Number	Yes	Yes *	
	Yes *	Yes	Yes *	
Population served	Yes *	Does not include faculty or aggregates for graduates and undergraduates separately	Yes *	
Collections				
No. of volumes at end of year	Yes *	Yes	---	Except for school with centralized libraries
No. of volumes added during the year	Yes *	Yes	---	
No. of periodicals received	---	Yes	---	
Circulation				
Total circulation	Yes *	---	---	
Juvenile circulation (where applicable)	Yes *	Not applicable	---	

Source: 1962 U.S. Department of Health, Education and Welfare, Office of Education, Statistics of Public Library Systems in Cities with Populations of 100,000 or more and Fiscal year 1962 and Statistics of Public Library Systems in Cities with Populations of 50,000 to 99,999, Fiscal year 1962.

1966 American Library Association, Library Statistics of Colleges and Universities 1965-66 Institutional Data.

1961 U.S. Department of Health, Education, and Welfare, Office of Education, Statistics of Public School Libraries, 1960-61.

Table I (Continued)

Personnel			
Total (in full-time equivalents)	Yes *	Yes	---
Professional personnel (in full-time equivalents)	Yes *	Yes	--- Except for school districts enrolled 150 pupils or more
Non-professional staff	---	Yes	---
No. of student hours	---	Yes	---
Operating Expenditures			
Total Institutional (where applicable)	Not applicable	---	---
Total Library (excludes capital outlay)	Yes *	Yes	--- Except for school districts enrolled 150 pupils or more
Salaries (professional)		Yes	---
Wages	---	Yes	---
Student Salaries	---	---	---
Books	Yes *	Reported with periodicals and other printed materials	Yes * (Includes pamphlets)

Table I (continued)

Periodicals	Yes	Reported Together	Yes
Other Printed Material	Yes		
Audiovisual	Yes	---	Yes
Binding	Yes	Yes	Yes
Capital Outlay for Building Sites and Construction	Yes	---	---
Income (where appropriate)	Only for 1962	Not applicable	---
Local	Yes	Not applicable	---
State	---	Not applicable	---
Federal	---	Not applicable	---
Endowment Income	---	Not applicable	---
Private Contributions	---	Not applicable	---

them. Some isolated studies do exist; for example, one reporting the salaries of state librarians.

Part III. On Rationality in a Data System

There is a tendency in the planning of a data gathering system to impute virtue to sheer proliferation of quantity. Since ignorance is to be deplored it would seem that the larger the amount of information assembled the greater the net contribution of the activity. In fact, this is certainly not the case. The most obvious oversight in such a point of view is that information gathering is expensive and that money devoted to the collection of trivial or largely irrelevant materials might have been put to better use elsewhere in the nation's libraries.

But perhaps even more important is the fact that, in information gathering, sheer quantity often defeats its own purpose. As no one knows better than the librarians, many disciplines are already struggling with a torrential flood of information which becomes increasingly more difficult to digest and to utilize effectively. Redundant materials are undesirable not merely because they are wasteful, but also because they clog the information system and render it more difficult to make use of the items that are pertinent.

All of this argues that an ideal information system must police itself and must ration severely the amount of data collection which it permits itself. Once this is recognized it becomes important to avoid, insofar as possible, capriciousness in the decisions as to the priority rankings of alternative classes of statistical data. Extensive experience confirms that sheer intuitive judgement in these matters is likely to lead to the acquisition of statistical matters which on their face seem "nice to know" but which on closer examination have very little bearing on planning or policy, while it characteristically fails to provide materials whose importance is not obvious but which turn out to have a crucial role in a rational decision process. It therefore becomes imperative that there be a more systematic approach to the matter of priorities.

In undertaking a first approximation to such an approach or, rather, just the presentation of a few notions on which such an approach may be based, we proceed in two steps. We ask first who might use the information, and second in what ways they might employ it. Only by going through these steps can one be sure that the materials chosen for collection are really needed and that nothing crucial has been omitted. We are painfully aware that some of our

examples will only illustrate our own lack of professional experience in library operation. Yet the reader will recognize that these specifics are of no importance at this stage of the discussion, and that he can correct any such shortcomings by substituting in his own mind more appropriate examples.

It would seem to us that library data serve primarily four major classes of users. First and most obvious, is the librarian himself who can use them to learn from the experience of others, to anticipate prospective developments which may have not yet manifested themselves at his own institution, to evaluate alternative operating policies available to him and to anticipate, with the aid of calculated trends, future needs and obligations on whose basis he can make his economic plans.

Second, the data will serve those whose professional concern is not the individual library institution but who deals rather with inter-library matters. Those who work in the relevant portions of the U.S. Office of Education, in the professional associations, in any consortia of libraries fall into this second category.

Third, there are the institutional users of library services who must plan for a scale of library operation commensurate with the requirements of their own prospective activities. Public school systems,

colleges and universities and business firms who need libraries, for example, in their research and development constitute the most obvious cases in point.

The final major class of groups and individuals who may be expected, directly or indirectly, to make use of library statistics, are those who provide the funds on which the libraries operate. They can, obviously, use the figures to anticipate the amount of their financial obligations and the different amounts and qualities of library service which can be expected to correspond to alternative budget levels.

One can no doubt think of other users of library statistics--students who may employ them in dissertations, publishers who may use them in planning the size of a printing, etc. But it would seem that these cases are largely peripheral and that a body of statistics which meets adequately the major needs of our four main sets of users would provide all that can reasonably be expected.

Perhaps even more important, and certainly less obvious, are the specific purposes to be served by the data for each of these groups. In an ultimate sense, a piece of information is pertinent only if it can potentially lead to a modified course of action. For example, suppose some figure indicates that a certain type of cost is disturbingly high or that a certain minority group is declining in its

use of the library. These figures may be disturbing, but if no one has any idea as to what can be done to change matters the information is not useful. It may satisfy what Veblen termed (with approbation) our "idle curiousity," e.g., it may represent an interesting piece of pure research, but it has served no purpose as part of the body of information that is needed for applied activity.

To be useful in this sense information should be expressed in conditional terms and should be expressed in terms of the relevant alternatives. For example, suppose in an area measuring four square miles one is considering whether to establish two public libraries each contain x volumes or to build a single larger library containing $2x$ volumes, and that the costs corresponding to the two proposals were very similar. The first arrangement offers the advantage of greater proximity to the borrower while the latter, because it may to some extent be able to avoid duplication, is likely to offer a greater variety of books. In that case, it would be extremely useful to know on the basis of experience in similar areas which of the two sorts of arrangement attracts the larger volume of activity, i.e., which is likely to produce the larger circulation.

This illustration suggests two general conclusions:

a) to determine the range of data that are relevant one must first list explicitly the range of alternatives available to the decision maker and the criteria on which he wishes to choose among them.

b) wherever possible, data should be provided in conditional form, associating them with values of related variables. That is, one should provide in accord with the preceding example, not just a time series giving circulation per library over a number of years, but circulation related to number of volumes stocked and size of population served.

To illustrate the sort of possibilities that should be included in a list of alternatives we provide in Appendix A a first attempt at such a listing. Since it has not been prepared by professional librarians such a listing cannot pretend to have taken into account all the pertinent issues, nor is there any reason to believe that everything it includes is really relevant. Its purpose is merely to serve as an example of what the appearance of such a list might be, and that is the only respect in which it is intended to be taken seriously.

For each of the alternatives in the list it would be desirable, where they are available, to collect data on costs, associated personnel requirements, and measures of library use, including circulation, number of persons served, etc.

Of course, some of the pertinent figures will prove unobtainable simply because they are not currently recorded by anyone or because they go beyond the range of current experience. We cannot really report

the costs of a "fully automated" library of some given specifications if such a library has never been constructed and operated. However, a list of data constructed on the principles recommended would offer the advantage of coming as close as possible to the data set pertinent to the available decision alternatives. In addition, it would identify automatically the lacunae in the available statistics and make it easier to judge whether direct experimentation or some other procedure is worth undertaking as a means to fill these gaps.

Before concluding this section it is appropriate to include a few remarks about one class of information that is frequently asked for--the cost associated with various distinct segments of library operation, e.g., the cost of circulation, the cost of reference information, etc. There are two grounds on which the economist discourages interest in such figures. First, he maintains that they are unobtainable in principle and that the figures that purport to represent these magnitudes are generally arbitrary and misleading; second, he maintains that even if the data were available they would be of little if any legitimate use.

The data are unobtainable because there are large elements of common and inseparable cost in any library operation. Suppose a small local library tries to allocate costs between circulation among adults and children. A variety of outlays ranging from administration

to air conditioning of the building are incurred in common by the two branches of activity. Should the costs be assigned in proportion to the number of volumes circulated? The number of persons served? The number (or the cost)of new volumes purchased by each? Clearly the choice is arbitrary and the answer will reflect no more than the convention adopted as the basis for this allocation process.

Moreover suppose it were found, e.g., that the cost of circulation per volume were 29 cents and the cost of cataloging per volume were $3.46. What conceivable action could be taken on the basis of this item of intelligence? Surely this is typical of the sort of information that tends to clog the information process and makes it more difficult to get at the really pertinent data.

Part IV. Some Recommendations

A. General Proposals

The main proposal offered by this report is that consideration be given to the publication of an annual compendium of library statistics. There can be little doubt about the value of such a volume. Libraries are a most important element in the nation's educational, cultural and research equipment. It is incredible that so little should be known about the economics of their operation and that so little reliable information should be available as a basis for planning and

decision making. Few activities of comparable national importance are so poorly documented. With growing federal support for the libraries, it seems reasonable that a small additional amount be provided to finance the information flow that might increase greatly the efficiency with which these funds and, indeed, the entire set of library resources, are utilized.

It is recommended that the publication be organized on an annual basis because otherwise information gathering efforts are likely to prove sporadic, with obsolescence rapidly overtaking some of the figures and large gaps rendering many of the time series all but unusable. Moreover, an annual publication can be prepared by a permanent staff whose efficiency is enhanced by specialization and experience and whose permanence obviates the heavy costs that are incurred each time a research group is assembled anew.

An annual publication also offers the advantage of standardization. When data for different years are compiled by different sets of personnel the specifications of the figures are very likely to change. The comparability and hence the usefulness of the statistics are thereby likely to be reduced severely.

Finally an annual, or at any rate, a periodic publication is more useful because with the passage of time it becomes more familiar to its users. They are more likely to become aware of its

existence with the passage of time and, they are more likely to become better acquainted with its contents, its uses and limitations.

The annual compendium of library statistics should, of course, include a detailed "user's guide" which describes changes in sample and coverage, modifications in definition, etc., as well as providing standard information in the construction of the series, the definition of the various statistical categories and the other types of information needed for effective utilization of the data.

Our second major suggestion is that, wherever possible, data be collected on the basis of a constant sample. That is, rather than attempting to collect information for all libraries of a given class, regular time series be based, where appropriate, on figures collected from a fixed set of institutions the accuracy of whose responses can be relied upon. It is tempting to try to construct a universal sample which reports information for the entire population . However, this has a number of disadvantages which are not always recognized. Aside from the far greater cost in terms of effort and money that is involved in the processing of large numbers of responses and in following up institutions which fail to respond on a first request, a quasi-universal sample is likely to suffer from deficiencies in accuracy and comparability.

Where information is received from very many respondents one is not likely to be acquainted with the preponderance of the respondents, the quality of their records and the care with which their replies are formulated. The accuracy of such figures may therefore be a cut below that of data obtained from a smaller, carefully-constructed sample. Moreover, experience shows that it is virtually impossible to obtain complete returns from any population of potential respondents. The attempt to obtain a universal response is therefore likely to produce a sample, but one whose composition has been determined fortuitously rather than by deliberate design. The self-selected sample is likely to be biased, in that those who fail to respond will not be a random selection of the whole but is likely to be composed heavily of institutions that are poorly financed and poorly operated.

Even more serious is the fact that the sample will fluctuate in size and in composition, not only as because there will be variation in the set of non-respondents, but also because some new libraries open and others may perhaps close. All of this makes it very difficult to utilize the resulting time series for any analytic purposes.

We therefore propose that at least the initial data collection effort for the recommended statistical annual be construction on the basis of samples that are limited in size but carefully designed, samples corresponding to each of the various classes of institution--the public libraries, the school libraries, the college and university libraries, etc. This arrangement will keep the magnitude of the

initial effort within reasonable bounds and yet can provide usable and indeed rather reliable sets of materials. Later, if it proves desirable, it will be possible to increase the sizes of the samples. Here it is necessary to provide a transition period during which the behavior of the smaller and larger samples are compared and with whose aid the earlier and the later series can, as it were, be "spliced." An occasional general survey can also serve as the basis for an extrapolation of the sample data to the entire universe of libraries.

As a third general suggestion, it is proposed that there be collected data necessary for the rational planning of capital construction. The most obvious approach to this issue is a survey in which respondents are simply asked about the adequacy of their space ("much more needed "more needed," "adequate," "excess"). The resulting information might then be aggregated by weighting each answer in terms of the number of volumes in the respondents library.

A more useful set of statistics might be obtained by determining for each respondent library the absolute amount and percentage of his shelf space that is unfilled, and a comparison of these data with the corresponding figures for the preceding year.

The respondent might also be asked to estimate the maximum circulation per year that his current facilities would enable him to

sustain, and this could be compared with actual circulation figures for the current year and for some preceding year. If it is decided that circulation is an unfortunate index for this purpose, some other measure of use might be substituted. The important point is not the detail of this proposed procedure. What is essential is recognition of the significance of some measure of current capacity and the rate at which reserve capacity is used up. For without such information, rational planning for the major expenditures involved in the provision of library capital becomes all but impossible.

Finally, we offer several general suggestions which we consider highly desirable though they are obviously of less critical significance.

a) In the publication of information it might be desirable to provide less detail on the breakdown of salary statistics and other specific items. For very minor subcategories, it is difficult to enforce uniformity in the definitions of the basic data and the resulting figures are likely to prove unreliable and difficult to interpret.

b) In light of the very great interest in automation in the libraries, it would be appropriate to begin the collection of systematic information on the subject. At least to begin with, each respondent should be asked to report his expenditures on automating equipment, both in absolute dollars

and as a percent of his total equipment purchase expenditure. The figures should also be broken down by function (circulation, reference, etc.).

c) Materials are one of the most important categories of library outlay, yet no satisfactory price index seems to be available for books and periodicals. The index currently published in the Digest of Educational Statistics is somewhat misleading and therefore probably should be abandoned. This index measures the average price of new card cover titles. Since it is unweighted, the rising of number of high-price low-volume "art" books published in parts is a serious upward bias to the rates of increase of prices. We recommend that the construction of a fixed weighted price index for books and periodicals should be undertaken. A weighted index should be calculated from data on current cost per book and per periodical, with weights provided by the value of purchases in the base period, if this proves practical. It may prove that further investigation will indicate that the only practical solution to this problem is to construct a weighted index by type of book (e.g., University Press books) and leave it to the individual libraries to weight these indexes in accordance with their own purchasing patterns.

B. Some Specific Suggestions Relating to Particular Types of Library

Since for all practical purposes, usable statistics are currently available only for public school and college and university libraries, specific

suggestions can be offered only for these categories of institution. We begin with a few comments on the first of these categories.

1. Public libraries

a. It is recommended that both population served[10] (as derived from U.S. Census figures) and registered borrowers be recorded, in order to measure not only the magnitude potential of library services but also the extent to which they are used. A count of registered borrowers will not measure the number of people who come into the library for on-the-spot use, or the number of people using a single card, but, the count over time (as compared to population served) should offer some indication of trends in the effectiveness of the libraries.

b. Circulation figures should continue to be collected despite the ambiguity of their definition, because they are probably the best available measure of the volume of service provided by the public libraries.

c. Income figures should in the future be recorded (as was done for fiscal 1962). For without these income figures, it will be impossible to estimate the magnitude of current and prospective deficits and hence financial planning will be severely handicapped.

---

[10] Population served should be reported only for years close to the census years unless some attempt is made to bring the figures up to date for intermediate years.

2. College and university libraries

As was indicated earlier in this report, college and university libraries provide the most reliable and most comprehensive data for any category of libraries. The very quality of these figures makes it possible to suggest a number of ways in which their value might be increased even further.

a. It is hoped that aggregates for the U.S. will continue to be provided but, in order to facilitate comparisons with homogeneous groupings it is recommended that the data also be aggregated by type of institution (university, liberal arts college, teachers college, junior college).

b. It is recommended that full and part time faculty also be listed and taken to comprise part of the "population served." Ideally, these would be measured in terms of full time equivalents so that two half-time faculty members would be counted as a single faculty member.

c. A distinction should be drawn between undergraduate and graduate students.

d. In order to provide some measure of activity or "output" of the libraries, circulation figures should be reported. It is to be hoped that a standard unit can be defined so that comparability of data throughout the nation can be maintained. Should it prove impossible to define such a unit (as a result of wide variation in loan periods "reserve" policies, etc.), individual circulation figures which are at least internally consistent can be obtained. Hence, the comparison of

aggregates over time would still provide a legitimate indication of trends in library activity. That is, even if the absolute figures were not capable of any reasonable interpretation because of the heterogeniety of their components, the time series would still serve as a reasonable index of the trend of library activity over time.

e. It would also be desirable to record library activity in terms of "library-users": a turnstile count of the total number of people that enter the library daily or during the year (regardless of variability in the length of the "library day").

f. Finally, and perhaps most important, an historical series for college and university libraries should be constructed. The figures should go back to 1947 or earlier. The raw data are available. All that is required is the selection of a representative sample which could be inflated to obtain estimates of the figures for all American college and university libraries. This was essentially the procedure employed by us on a limited scale in our study of the economics of libraries. A long term series of this sort could be obtained with very little cost outlay.

C. The Organization of Federal Efforts to Gather Library Statistics

Since the reorganization of all education data-gathering efforts into the National Center for Educational Statistics has only recently occurred, it is at present impossible to evaluate the full impact of this move on Library Statistics. The National Center is now in charge of compiling statistics on the three kinds of libraries formerly serviced by program specialists: the public libraries, school libraries, and college and university libraries. During this transitional stage the National Center has not yet been able to set up a smoothly functioning organization in the library area.

The imperfection of the activities in this area is clearly illustrated by the extremely long lags in the availability of current data. The latest published Federal statistics for public libraries, public school libraries, and college and university libraries are those for the fiscal years 1962, 1963 and 1964 respectively. Some of these data are now six years old. It appears that excessive effort has been devoted to the publication of large amounts of detail for college and university libraries, --and not enough attention has been paid to the provision of basic data on as current a basis as possible.

If the National Center for Educational Statistics is able to mobilize sufficient staff to provide timely publication of at least the same kinds of statistics as have been previously published, with the

corrections and additions we have recommended above, then the centralization of Federal data-gathering for libraries should yield significant improvements in the quality of the data published. Thus we endorse the concept of centralized data collection as potentially the most efficient method of providing data of good quality. However, centralization alone is no panacea and must be accompanied by adequate staff.

The reorganized National Center for Educational Statistics must obviously be given a period of a year or two in which to organize its efforts and demonstrate what it can accomplish. There is every reason to hope that its efforts will prove productive. However, should it turn out after such an internal that there has been no material improvement in the quality and currency of the available information alternative means of providing the data will have to be explored.

Two possibilities will then merit consideration. It may prove desirable to have data collection sponsored by the National Library Association much as the American Association of University Professors has undertaken the task of providing current figures on faculty compensations, a piece of work which it has been performing at any reasonable cost for some two decades. A second alternative is to contract for the work to be done by an outside agency -- a research organization such as the Brookings Institution, or a firm

specializing in economic research. This would indoubtedly require some federal support but our own experience indicates that the data are collected on a sampling basis, they can be acquired and organized at surprisingly little cost.

It should be emphasized again that consideration of these alternatives at this point is certainly premature. There is every reason to be sanguine about the prospect for the work of the National Center. Our only objective in the preceding paragraphs was to suggest that alternatives do exist, and to indicate to whom one might turn in the event that it becomes necessary to explore them.

The specific recommendation outlined in this report encompass the areas in which we feel that effort will be most useful. It is hoped that a significant advance might be made in some of these areas within the next year; such an effort will indicate that the National Center is on the road for fulfilling its promise.

Part V. Concluding Comment

This survey concludes as it began--with the observation that the issues involved in the collection of library statistics have been explored carefully and intelligently. But the data themselves are

in a deplorable state. Only for the college and university libraries is any sort of systematic and recent information available. Given the importance of the libraries to the nation and the increase in efficiency of their operation which might be made possible by better information, the relatively small outlay which would be required for the provision of systematic statistical information seems clearly merited.

APPENDIX: Preliminary Classification of Some Library Activities and Some Corresponding Decision Alternatives

Administrative Services

Policy Determination

Planning, review and evaluation

Budget

Accounting, bookkeeping and auditing

Clerical, mechanized or hiring of outside personnel and/or machines

Personnel coordination and supervision

Recruiting

Training

Classification

Service rating and promotion

Payroll and fringe benefits

Last can be mechanized--others require professionals

Building and Business management

Space planning

Can be handled by outside professionals or planned from experience, suggestions, professional articles, etc.

Operation and maintenance of physical plant

Janitorial Function

Supplies

Tabulation and ordering can be mechanized rather than clerical

Duplicating and Photography

Varying degrees of mechanization, but all require operator

Shipping and delivery

Can use outside agency rather than own personnel

Routing and communications

Own personnel required

Public Relations

Publicity - promotion

Exhibits

Tours

Could be handled by pooling or by using central agency for whole system, but need own for individual libraries.

Annual reports and statistical information could be centralized or handled by outside professionals.

Decision Areas

Readers' Services

Central and Extension services

Branches

Deposits

Stations

Schools

Bookmobiles

System activities

Circulation

Registration

Can be computerized or handled at central point

Charging

Can use photo-charge or computer-base system to transfer to magnetic tapes--old system of hand stamp (photo-charge--about 6 different systems)

Overdues

Magnetic tapes can be fed to computer to give list of overdues and print notification cards

Reserves

Program computer to note reserve, print card or staff must check files, send notice

Shelving

Open or closed shelves must mark and file

Stack maintenance

Usually done by staff--difficult to do otherwise

Interlibrary loan

Request--must standardize catalogs--can use computer

Photocopying

Use of very small reduction--standard reduction--either same day or wait services

Reference

Direct service and reader assistance

Employ professionals--Computer-call supplies references

Mail--telephone

Professionals--can also use computer-call

Interlibrary

Reference

Bibliographies

Selected lists

Annotated

Indexes

Literature searches

Abstracting

Translating

All can be done by professionals, but computer is ideal for all above situations.

Readers' Advisory services

Library Instruction

Individuals

Groups, special clientele

Adult education

Need professional guidance--centralized library would allow pooling.

Decision Areas

Technical Services

Acquisition and withdrawal

Searching of titles

Ordering

Accession

Record keeping

Routing

Gifts and exchanges

Weeding

Disposing

All functions are ideally suited for computer tape storage. Records could be used for decision-making by personnel. Use of clerical help is standard method

Classifying and cataloging

Verifying

Cataloging

Shelf-listing

Catalog maintenance

Call numbering

Standard process uses few machines, many people. Some cataloging devices exist. Computer could be used for most functions.

Preparation

Bookmarking

Pocketing

Jacketing

Require personnel

Binding

Rebinding

Mending

Repairing

Generally, work is farmed out - central areas could handle work for many.